

分布式虚拟化存储在公安专科学校中的应用

朱康林

(上海公安高等专科学校 上海 200137)

摘 要 随着云计算、大数据等新技术和应用的不断涌现,传统 SAN 存储在扩展性等方面暴露出不足之处。虽然依照传统存储的思路构建了大容量+高性能的存储架构,但是在数据量大幅增加、云计算和互联网应用逐渐增多的情况下,公安专科学校需要一套新的具有高性能、易扩展、易管理等特性的存储架构。vSAN 通过软件的不断更新升级,能支持越来越多的功能,性能也得到不断提升,其产品发展进程已经完全改变了原来以硬件为主的存储架构演进,随着数据镜像复制、虚拟机双机热备、容灾等高级功能的推出,其完全突破了传统 IT 架构,大幅度提高了系统的性能和安全性,IT 管理、数据管理等都将面临变革。上海公安高等专科学校是国内首个成功部署并上线运行 VMware vSAN 虚拟化分布式存储的单位,它在 VMware vSAN 项目实施过程中遇到的问题以及应对之策对于其他用户的分布式存储实践和 VMware 公司的产品改进都具有积极意义。

关键词 分布式存储,虚拟化存储,VMware,vSAN,云计算

中图分类号 TP311 文献标识码 A

Application of Distributed Virtualized Storage in Public Security College

ZHU Kang-lin

(Shanghai Police College, Shanghai 200137, China)

Abstract With the development of cloud computing and big data, traditional SAN based on storage technology cannot meet the new need. With the emerging need of huge storage and applications on web, public security college still needs a new storage framework. Through the refresh of software and new function, the performance has been improved. The development of the product has changed the evolving based on hardware. In this paper, we introduced the work using vSAN and applications on Shanghai Police College.

Keywords Distributed storage, Virtualized storage, VMware, vSAN, Cloud computing

1 引言

IDC 的统计报告显示,到 2020 年,全球数据总量将达到 40ZB。数据量呈线性快速增加,加大了对存储的压力。在这种情况下,传统的 SAN 存储还能继续发挥它的效能吗?随着云计算、大数据等新兴技术和应用的快速兴起,虚拟化、分布式等技术在存储应用中发挥着越来越重要的作用。一方面是数据增长带来的挑战,另一方面是存储技术本身的变革。上述诸多方面的压力反映到高校,具体表现为:高校迫切需要更高性能、更易扩展和更易管理的存储系统,以确保教学活动的正常开展。

2014 年 3 月,VMware 公司正式发布了其软件定义存储。与传统 SAN 存储最大的不同表现在,它直接运行在服务器上,采用分布式集群架构,依靠闪存实现应用加速,并且与 VMware vSphere 虚拟化软件高度集成。2014 年,上海公安高等专科学校成了中国第一个 VMware vSAN 成功上线运行的用户;2015 年 3 月,升级 VMware vSAN 到第二个版本 VMware vSAN6.0,同时上线了完整的 VDP 备份系统,并进行了相关的测试。

传统的存储架构主要包括 3 种:直连式存储(DAS)、存储

区域网络(SAN)和网络接入存储(NAS)。这 3 种存储有各自适用的场景,比如 SAN 通常应用于对性能有较高要求的数据库类应用,而 NAS 主要用于存储文件和非结构化的数据。

从网络存储概念的出现到现在的几十年中,SAN 存储技术已经发展得非常成熟,一直被认为是主流的也是应用最广泛的一种存储方式。在企业级用户的眼中,SAN 存储甚至成了高性能、高可用、高可扩展存储的代名词。但是,SAN 存储也有其局限性,比如成本和复杂性高。随着云计算、大数据、移动化、社交化等新技术和应用的不断涌现,人们需要随时随地快速地访问企业内外的数据,这时传统的基于纵向扩展模式的 SAN 存储在扩展性方面就显得捉襟见肘。

传统存储方式的最大问题是不具有横向扩展的能力,也就是它只能在一个主机下面进行扩展,受主机限制,其扩展只能是纵向的,且有限度。随着云计算、大数据等应用的兴起,数据存储量呈几何级数增长,因扩展方式的限制,传统存储方式无论具有多高的性能也无法满足系统增长的需要。如果用户通过不断购买新的存储设备来实现存储容量的升级,其高昂的成本将是用户无法承受的。

早在 2000 年左右,存储虚拟化技术就已经诞生。存储虚拟化就是将原来分散的、异构的存储资源统一整合成一个虚

拟的存储资源池,对其进行统一管理、调配和使用。存储虚拟化的优势十分明显,比如可以整合异构存储,提高存储的利用率,降低成本,提高存储的灵活性等。不过,我国的行业用户在实际部署存储虚拟化方面并不如想像中那样广泛。一方面,存储硬件厂商的技术封闭性,使得存储整合还需要一个过程;另一方面,缺少杀手级的应用迫使硬件厂商走开放的道路也是存储虚拟化现阶段不能快速普及的主要原因之一。不过,在服务器虚拟化技术快速普及的基础上,许多用户已经开始尝试存储虚拟化。随着存储虚拟化技术不断走向成熟,存储虚拟化具备的优势正逐步展现出来。以服务器虚拟化起家的 VMware 公司,在 2014 年相继发布了虚拟化网络、虚拟化存储产品。与传统 SAN 存储相比,VMware vSAN 虚拟存储区域网络技术显示出了以下优势。

(1) VMware vSAN 可以有效降低存储的采购和使用成本。VMware vSAN 使用标准化、通用的 x86 服务器作为存储器,这是具有创新性的。此举使用户摆脱了长期以来存储产品采购严重依赖存储厂商的情况,将大大降低存储的采购成本。此外,服务器在达到使用寿命后,可以通过自然更新提升系统的性能,而不是像以前那样必须成套更换存储系统,从而降低了存储的使用成本。

(2) VMware vSAN 的数据安全性更高。传统 SAN 存储的价格通常较高,而且为了保证传统 SAN 这种集中式存储的可靠性,就需要购买多套存储并建立冗余网络等,这样不仅成本高而且管理复杂。但是,采用 VMware vSAN 这样的分布式存储技术,当集群中的服务器出现故障(即使硬盘损坏)时,都没有太大影响,只要在方案设计时考虑到服务器、磁盘、网络等的冗余即可保证整体的可靠性。

(3) VMware vSAN 的管理更简单。传统的存储管理是各自为政的,并且不同的存储有不同的管理软件,非常复杂,管理员的学习成本也很高。VMware vSAN 极大地简化了存储管理,因为它可以与已有的 VMware vSphere 服务器虚拟化管理整合在一起。管理员之前如果使用过 VMware vSphere 服务器虚拟化技术,那么管理和使用 VMware vSAN 则易如反掌。

(4) VMware vSAN 的性能并不逊于传统 SAN。VMware vSAN 通过两个有效的技术手段来提高存储性能。1)它采用分布式存储技术,不仅提高了数据的安全性,而且由于分布式本身的优势,使得数据读写的性能成倍提高,以后随着节点数量的增加,整体的性能也可以实现线性增长;2)通过使用 SSD(固态硬盘),使得性能得到显著提升。

(5) VMware vSAN 原生支持分布式系统,为云计算和大数据提供理想的硬件支持。分布式存储系统的特点是可以将数据分散存储在多台独立的设备上。传统的 SAN 存储系统采用集中的存储服务器存放所有数据,因此存储服务器成为系统性能、可靠性和安全性的瓶颈,不能满足大规模存储应用的需要。而分布式存储系统采用横向扩展的系统结构,利用多台存储服务器分担存储负荷,并利用位置服务器定位存储信息,不仅可以提高系统的可靠性、可用性和效率,而且还方便扩展。如今,分布式存储已经成为广大用户关注的焦点,尤其是在大规模的云计算系统中,已经是一种理想的存储解决方案。

2 现状分析

数据中心是云计算环境下分布式存储的基础,它是构建云的最基本的数据存储单位。数据中心网络利用互联网将不同的数据中心连接起来构建成整个云的存储设施,为云的使用者提供数据服务^[1,7-9]。上海公安高等专科学校是一所中等规模的高职、高专院校,其信息化系统的建设主要以满足校内各类职前职后培训和学习的需要为目的。从具体应用需求来看,校内的应用并发使用人数一般在 2000 以下。2006 年,学校确定了“一套高性能存储+一套大容量存储”的数据分层存储顶层设计方案,即用中高端光纤 SAN 存储(EMC Celerra NS960)来解决数据性能的问题,满足核心业务系统的应用需要,辅以 NAS 存储(EMC Isilon)解决数据存储容量问题,满足视频、云盘等大容量数据存储的需要。此后,学校一直沿着这一思路不断扩展和升级其存储系统,满足了各类新建业务系统的存储需要,此外又结合服务器虚拟化技术,将信息系统的部署和管理效率提高了 10 倍以上。但是,随着新应用和新技术的不断涌现,当初建立时认为比较完善的存储框架的不足逐步显现,归纳起来有以下方面。

(1) 学校数字图书馆早期购入的数字资源所采用的开发技术老旧,无法迁移到新购买的存储系统中。以“知网系统”为例,它每年都要新增 4TB 以上的存储空间,而学校一直依靠不断更换更大存储容量的服务器来满足不断扩张的存储容量需求,但每次更换新的存储设备都需要 2~3 个月,人力成本高昂,这种方法不具有可持续性。

(2) 学校在教学应用的环境中长期使用的无盘工作站,由于产品自身的原因,无法提供高可用性,也无法采用虚拟服务器技术来支撑,一旦设备出现故障将导致应用无法正常运行,严重影响教学效果。

(3) 学校开始逐步将云计算、大数据等最新技术用于教学和学校的信息化建设,传统存储架构无法支撑云计算、大数据的需求。实践证明,云计算、大数据应用只有在分布式存储的环境中才能充分发挥其作用。

(4) 传统存储设备的扩充十分依赖厂商,导致产品后续升级和服务的成本很高。

(5) 目前,EMC Celerra NS960 的核心存储是学校最为重要的存储设备。该设备的可靠性虽然很高,但仍存在设备单点故障。一旦该设备发生故障,将严重影响学校的正常教学和业务开展。厂商提出的解决方案是再购买一套完全相同的存储作数据镜像,但成本过高,不具有可行性。

(6) 传统存储的架构方式是从下往上部署的形式,最先规划存储的使用量,然后配置 RAID 和划分逻辑卷,一旦配置划分完成并开始投入使用后,便难以再对其做更改,即使有办法也需要严谨的规划且会消耗大量的时间和资源。由于传统存储的部署方式是从下往上的,因此传统存储对于虚拟化环境、虚拟机是无法对其认知并进行管理的。对于传统存储而言,划分出逻辑卷以后,SLA 就被固定了,并且部署在上面的所有虚拟机将有着相同的 SLA。简单且粗暴的规划方式,不利于满足虚拟化多变、灵活的业务需求。

正是考虑到传统存储具有以上不足,上海公安高等专科学校为了适应云计算、大数据的需要,考虑采用存储虚拟化技术。VMware vSAN 虚拟化存储就是比较理想的选择。

当前主要有 3 种类型的数据中心网络:交换机中心网络、服务器中心网络以及混合网络。以交换机为中心的结构利用

交换机将数据中心上的各个服务器连接起来,由交换机负责数据包的路由转发,服务器只负责数据的存储和处理。在以服务器为中心的结构中,服务器除了承担数据的存储和处理外,还需要负责数据包的转发。在该结构中需要在服务器上安装多个网卡,然后通过网线把这些服务器连接起来。在这种结构中,不需要交换机等数据转发设备。混合结构是上述两种结构的一种混合结构,其中既有交换机,也需要部分具备数据转发功能的服务器。Greenberg 等人提出的 VL2^[3] 树型结构是一种改进的以交换机为中心的数据中心网络,其利用了 Clos 网络结构^[4],解决了传统树形结构中核心层交换机数量少而造成的大量数据转发带宽瓶颈问题。同时,在 VL2 中还实现了服务和应用与服务器地址的分离。在部署时,服务和应用可以部署在数据中心的任意服务器上,并通过名称加以识别;在服务和应用运行时,通过构建映射得到服务和应用所在服务器的地址,提高了资源的利用率。但是,VL2 依旧没有解决以交换机为中心的数据中心网络中服务器之间存在的链路容量有限、资源浪费以及扩展性问题。Libdeh 等人提出的 CamCube^[5] 数据中心网络结构是一种以服务器为中心的结构。在 CamCube 结构中,没有任何的交换机和路由设备,服务器间利用多个网卡相互连接,并由服务器完成数据包的路由转发。CamCube 不存在单点瓶颈,结构和线路连接都很简单,但是,路由转发功能将会大量占用服务器的计算资源,对服务器造成极大的负担,而且服务器所能装载的网卡数量有限,从而限制了 CamCube 的节点规模。Guo 等人在 2008 年提出的 DCdl^[6] 是一种混合结构的数据中心网络。其在设计过程中充分考虑到了链路的容错能力,采用了一种分层递归的结构。相较于传统服务器为中心的结构,DCdl 能够更好地聚合带宽,能够较好地处理数据密集型计算中一对多和多对多的通信要求。但是,DCcell 运行过程中会产生大量的链路冗余,而且其构造成本较高。

3 vSAN 虚拟化存储实践

3.1 内网建成虚拟化存储平台

确定采用 VMware vSAN 技术后,上海公安高等专科学校制定了技术方案,在内网环境(和互联网物理隔离),采用 1 台万兆交换机加上 5 台服务器组成集群,以及 2 台服务器用作管控和备份;于 2014 年 6 月 13 日完成系统部署,成功上线运行,成为国内首个部署运行的 vSAN 系统。系统的拓扑结构如图 1 所示。

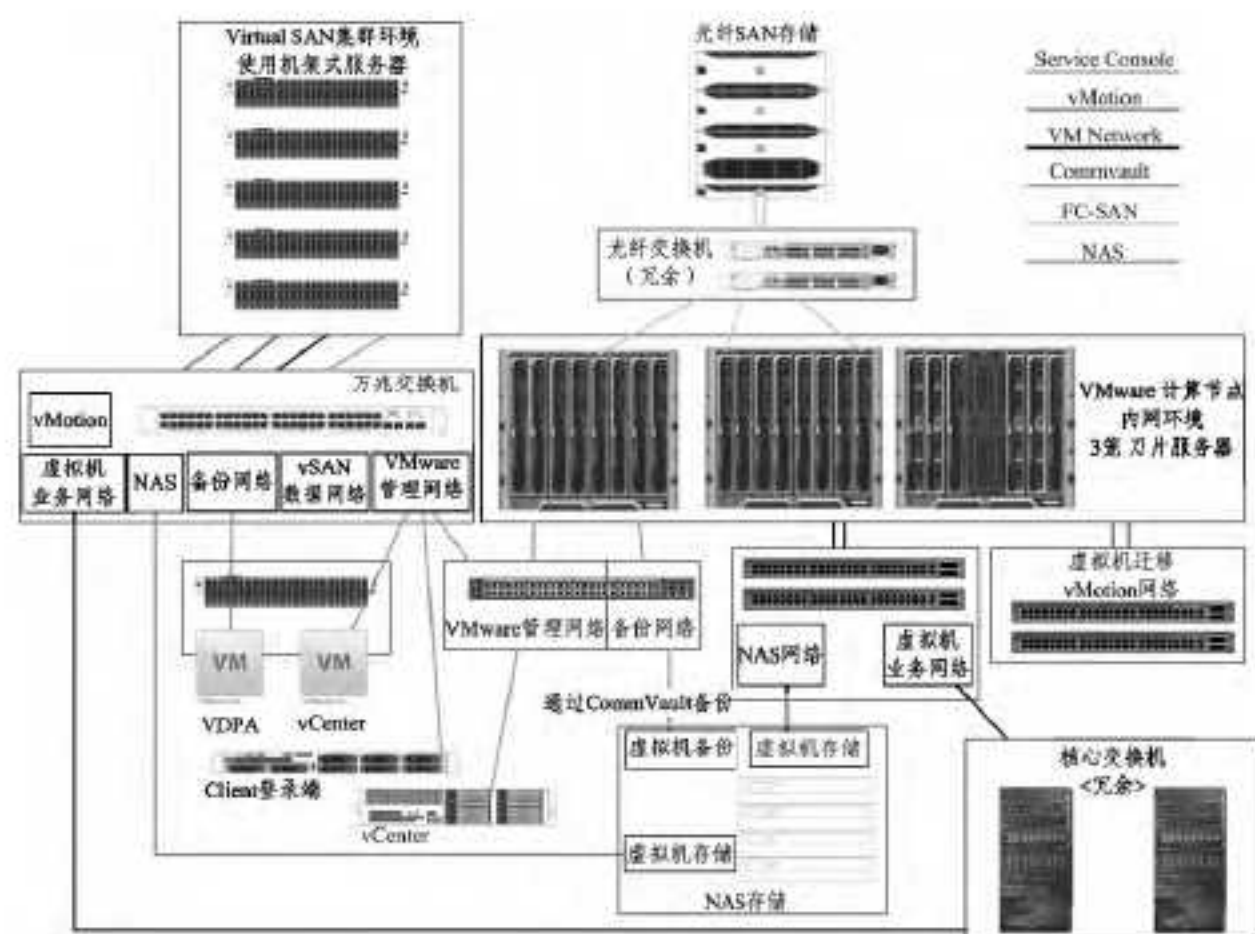


图 1 系统的拓扑结构

在首次建成的 VMware vSAN 虚拟化存储区域网络系统上部署了学校无盘工作站和云桌面系统,并对其性能和数据安全进行充分验证,达到了如下效果

(1) 在原有 VMware 架构体系内满足了大容量存储的需求。VMware vSAN 与 VMware vSphere 无缝结合, 可以大大提高存储的使用范围和管理效率, 并可保护现有的软硬件投资。

(2) 使用 VMware vSAN, 可以通过保留多份数据副本来提升数据的安全性。

(3)VMware vSAN 存储与成本相对较高的 IP SAN 集中存储的性能相当,完全可以替代中高端 IP SAN 存储。实际使用中,VMware vSAN 存储服务器配置的 SSD 和 SATA 硬盘互相呼应,当多个数字业务系统有多任务、高并发的突发流量产生时,在 IOPS 较大或者吞吐量增大的情况下,VMware vSAN 在理论上能够自动实现先将数据缓存到 SSD,然后再写入到 SATA 硬盘中,没有出现当多个检索系统服务器多任务高并发运行时,存储后端读写速度缓慢或减慢的情况。

(4)VMware vSAN 方便进行存储容量扩展。后续需要增加存储空间时,只需要增加普通的 x86 服务器,按照之前的配置要求加入到相应的存储服务器组即可完成扩容,无需停机或业务停顿,业务的连续性得到了充分保证。

(5) 采用 VMware vSAN 分布式存储后,可以不再依赖某一种或某几种存储,后续扩容可以避免高昂的服务成本和硬件扩容成本,只要考虑 x86 服务器的服务和硬件扩容成本即可。后续存储投入成本只是原来的一半,具有很高的性价比。对于公安院校和类似规模的高职高专的数据中心来说,VMware vSAN 部署方式简单、快捷,业务不会中断,更不会影响原有存储架构的性能。

VMware vSAN 分布式存储最大的优势在于可以横向扩展,而且不需要一次性投入非常高的成本购买昂贵的存储设备,最大程度地保护了用户的原有投资。这也是很多互联网、云计算的客户愿意选择分布式存储的重要原因之一。互联网、云计算等应用的特点是需要灵活的资源调配,能够满足大量用户的并发访问,以及具有高可用性的保障。学校这次部署 VMware vSAN,可以针对不同的虚拟机磁盘设置不同的存储策略,这种细粒度的策略设置可以更灵活地实现应用调优。传统存储由于无法识别 VMDK,因此不能针对单个虚拟机配置策略。上海公安高等专科学校希望利用 VMware vSAN 的独特之处为学校的互联网应用平台提供更好的技术支持。

3.2 外网建成虚拟化存储平台支撑下的完整私有云

学校外网面临建设大量信息化平台的需要,之前仍采用传统的存储架构,亟待升级。在内网部署虚拟化存储平台的基础上,学校规划了完整的在虚拟化存储平台支撑下的外网私有云架构,来支撑学校未来 5 年的信息化建设。2015 年 8 月,学校建成了最新的外网私有云架构,通过部署最新版本的 vSphere 6.0+vSAN 6.0+vDPA,性能、安全性、扩展性等都达到了预期的目标。

测试系统: Redhat Linux 6.5 x64

测试工具: FIO, 如表 1 所列。

表 1 测试表

读写 块大小	顺序	读写占比	测试偏向	求和项: 吞吐量 (kB)	求和项: IOPS	求和项: 响应延迟 (单位:μs)
4k	顺序	0%读 100%写	写	11416	2854	2160
		0%读 100%写	汇总	11416	2854	2160
		100%读 0%写	读	204268	51066	1848
		100%读 0%写	汇总	204268	51066	1848
		70%读 30%写	读	109872	27468	2512
		70%读 30%写	写	47070	11767	2800
	随机	70%读 30%写	汇总	156942	39235	5312
		0%读 100%写	写	53431	13357	12864
		0%读 100%写	汇总	53431	13357	12864
		100%读 0%写	读	102257	25564	6880
		100%读 0%写	汇总	102257	25564	6880
		70%读 30%写	读	22911	5727	8032
64k	顺序	70%读 30%写	写	9839	2459	15000
		70%读 30%写	汇总	32750	8168	23032
		0%读 100%写	写	361303	5645	28000
		0%读 100%写	汇总	361303	5645	28000
		100%读 0%写	读	534811	8356	11840
		100%读 0%写	汇总	534811	8356	11840
	随机	70%读 30%写	读	495775	7746	6176
		70%读 30%写	写	213418	3334	50000
		70%读 30%写	汇总	709193	11080	56176
		0%读 100%写	写	182386	2849	32000
		0%读 100%写	汇总	182386	2849	32000
		100%读 0%写	读	781042	12203	9792
64k	顺序	100%读 0%写	汇总	781042	12203	9792
		70%读 30%写	读	529345	8271	11968
		70%读 30%写	写	227666	3557	21632
		70%读 30%写	汇总	757011	11828	33600
	随机	70%读 30%写	读	529345	8271	11968
		70%读 30%写	写	227666	3557	21632
		70%读 30%写	汇总	757011	11828	33600
		0%读 100%写	写	182386	2849	32000
		0%读 100%写	汇总	182386	2849	32000
		100%读 0%写	读	781042	12203	9792

测试环境: Windows 2008 R2 Enterprise Sp1

测试软件: IOmeter

测试目的: 测试虚拟机能够达到的 IOPS

数据块大小: 4kB

读写比: Read 100 / Write 0

IO(顺序、随机)占比分布: Sequential 100 / Random 0

任务数: 10 Jobs, 结果如图 2 所示。

Display	All Managers	24983.36	100000
Total I/Os per Second			>
Total MBs per Second (Decimal)	All Managers	102.25 MBPS (97.51 MiBPS)	1000
Average I/O Response Time (ms)	All Managers	0.3993	1
Maximum I/O Response Time (ms)	All Managers	15.7439	100
% CPU Utilization (total)	All Managers	19.86 %	100 %
Total Error Count	All Managers	0	0

图 2 测试结果 1

测试环境: Windows 2008 R2 Enterprise Sp1

测试软件: IOmeter

测试目的: 测试多台虚拟机同时运行时能够达到的吞吐量

4 台 VM 同时运行

Outstanding: 512

读写比: 随机 100%读, 0%写

数据块大小: 256k

VSAN 策略: FTT=1, Stripe=2

结果如图 3—图 6 所示。

Display	All Managers	2961.83	10000
Total I/Os per Second			>
Total MBs per Second (Decimal)	All Managers	776.43 MBPS (740.46 MiBPS)	1000
Average I/O Response Time (ms)	All Managers	1716.4231	10000
Maximum I/O Response Time (ms)	All Managers	4193.1983	10000
% CPU Utilization (total)	All Managers	4.16 %	10 %
Total Error Count	All Managers	0	0

图 3 测试结果 VM-1

Display	All Managers	1546.08	10000
Total I/Os per Second			>
Total MBs per Second (Decimal)	All Managers	405.30 MBPS (386.52 MiBPS)	1000
Average I/O Response Time (ms)	All Managers	3561.3321	10000
Maximum I/O Response Time (ms)	All Managers	5020.9901	10000
% CPU Utilization (total)	All Managers	2.86 %	10 %
Total Error Count	All Managers	0	0

图 4 测试结果 VM-2

Display	All Managers	2081.65	10000
Total I/Os per Second			>
Total MBs per Second (Decimal)	All Managers	545.59 MBPS (520.41 MiBPS)	1000
Average I/O Response Time (ms)	All Managers	246.0159	1000
Maximum I/O Response Time (ms)	All Managers	1486.8075	10000
% CPU Utilization (total)	All Managers	2.07 %	10 %
Total Error Count	All Managers	0	0

图 5 测试结果 VM-3

Display	All Managers	2193.61	10000
Total I/Os per Second			>
Total MBs per Second (Decimal)	All Managers	575.04 MBPS (548.40 MiBPS)	1000
Average I/O Response Time (ms)	All Managers	233.4416	1000
Maximum I/O Response Time (ms)	All Managers	1159.6396	10000
% CPU Utilization (total)	All Managers	2.26 %	10 %
Total Error Count	All Managers	0	0

图 6 测试结果 VM-4

可以看出,无论是 IOPS 还是吞吐量都有了大幅提升。

3.3 外网建成虚拟化存储平台支撑下的完整私有云

在 VMware vSAN 的部署环境中,会对硬件配置有一定的要求,也有兼容性方面的问题。不过,在 VMware vSAN 的最新版本中,兼容性问题将得到解决。在这个项目中,上海公安高等专科学校使用的是 Dell R720XD 服务器,所采用的 RAID 卡是 H310 mini,而在之后 VMware vSAN 取消了对 H310 RAID 卡的支持,VMware 公司对此予以免费更换 H710P 的 RAID 卡,以保证对新版本的支持。

VMware 推荐采用 H710P 的 RAID 卡。从性能来看,在传统的使用方法下(多块磁盘组成一个 RAID 阵列的情况),H710P 会比 H310 的读和写的速度都会快很多。因为 H710P RAID 卡带有高速缓存,而 H310 没有高速缓存,多块磁盘组成一个 RAID 阵列后,其读写效率会有一个很大的差距。H710P 具有较深的 IO 队列深度,而 H310 的队列深度非

常浅。当 IO 过多时, H310 会明显发生虚拟机、ESXi 主机卡顿的现象。H310 和 H710P 的另一个区别是, H310 支持 RAID 模式和 SAS 直通模式, H710P 只有 RAID 模式。如果要在 vSAN 环境下使用 H710 RAID 卡, 就必须将每一块磁盘都配置成一个独立的 RAID-0 逻辑卷, 这样才能让 VMware vSAN 识别并利用到这些磁盘。但带来的问题是, 如何在服务器不停机的情况下为服务器配置 RAID。在磁盘损坏后, 这块磁盘所建的单独的 RAID 会崩溃, 更换新的硬盘之后又不会有 Raid 重建的动作, 无法自动重新恢复这块磁盘应有的 RAID 信息。

给服务器配置 RAID 具有 2 种方法:

(1) 重启进入 RAID 菜单进行配置。这样的恢复操作容易对虚拟机的运行连续性造成影响。

(2) 使用 Dell 的服务器管理平台 OMSA (OpenManage Server Administrator) (可以监测并配置硬件信息), 需要在服务器 ESXi 系统内安装支持 ESXi 相对版本的 OMSA 插件, 并需要有一台安装了 OMSA 主程序的 Windows 的客户端, 通过 Web 远程登录 ESXi 服务器的 OMSA 服务。

这样配置可以做到即使有磁盘故障也不用重启服务器。虽然增加了网络带宽, 但对客户机系统启动的时间并没有太大的提升, 在客户端启动的瞬间读取磁盘的速度可能会直接影响启动的时间。上海公安高等专科学校进一步对 VMware vSAN 进行了测试。

测试结果, 15~20 个 worker, 32K~100% Read, 测试结果为 8000+ 的 IOPS, 260MB/s 左右的吞吐量; 15~20 个 worker, 4k~100% Read, 测试结果为 12000~16000 的 IOPS, 30~80MB/s 的吞吐量。

以上数据是 VMware vSAN 在做了两个条带的测试参数。参考这个数据, 上海公安高等专科学校认为性能是偏低的, 需要对 VMware vSAN 进行进一步的优化。

造成性能偏低的原因主要有以下几方面。

(1) 磁盘控制器的性能问题。环境最初部署时, 使用的是 DELL H310 RAID 卡 (该卡使用 LSI2008 芯片组, 具有 RAID0、1、5、10、50、NON-RAID 模式, 无电池单元, 无高速缓存)。使用这款 RAID 卡的目的是因为由于 RAID 卡是这款服务器唯一支持 NON-RAID 磁盘直通模式的阵列卡。在使用过程中发现了诸多问题, 后经多方调查、调试, 确认 DELL 的 H310 RAID 卡的任务队列深度非常浅, 因而导致磁盘整体 IO 堆积。呈现出现象就是, 一个很普通的任务, 却需要很长时间才能完成。发现问题后, 将这个型号的阵列卡换成 DELL 的 H710P RAID 卡。从理论上讲, 带有缓存的 RAID 卡性能应该是会有明显的提升。但 vSAN 环境能使用的磁盘模式只有两种, 直通模式和单块磁盘的 RAID0 模式 (必须关闭高速缓存中的“写缓存”功能, 可以开启读缓存)。因此在 vSAN 环境中, RAID 卡中高速缓存的效果不太理想。将 H310 更换成 H710P 的目的不在于高速缓存, 而在于队列深度。DELL 的 H710P 是这款服务器能够使用的最高端的磁盘控制器, 具有较高的队列深度和 1GB-DDR3 的高速缓存。

(2) SSD 磁盘的性能。在 VMware vSAN 环境中, SSD 是承载数据读写的缓存磁盘, 数据读写的命中率以及数据分配到哪个磁盘上都由 SSD 来完成。选用更高性能的 SSD 磁盘

可以提高 VMware vSAN 的整体性能。

(3) VMware vSAN 的磁盘策略。1) 条带策略。条带数量的增加可以提升磁盘的性能。VMware vSAN 的最高条带数量是 12。在 SSD 磁盘读取未命中的情况下, 条带数量成为关键。2) SSD 预留。可以预留一部分 SSD 空间, 并将一些重要的虚拟设置为常驻 SSD 的磁盘策略。在一些关键应用上或具有较高优先级的虚拟机上可以避免它们反复地将数据在 SSD 与 HDD 之间进行置换, 这样可以节省部分 IO 资源。3) 磁盘预留。在 VMware vSAN 环境中默认采用的磁盘策略是“1 个镜像+磁盘精简配置 Thin provision”, 而若想提升某台虚拟机的磁盘 IO 性能, 设置为“厚置备置零”是一个方法。因此这里的“磁盘预留”设置便是关于虚拟硬盘的空间预留策略。若能正确、合理地配置 vSAN 存储策略, 便可以更好地优化虚拟机。

(4) 网卡的 MTU (最大传输单元)。在 VMware vSAN 的环境中, MTU 的设置可能会对 VMware vSAN 的性能带来影响。

(5) VMware vSAN 重建的收敛时间。在日常维护中, 用户通常会对系统做一些常规的操作, 让计算资源、存储策略或主机进入维护模式。这些操作可能都会影响到 VMware vSAN, 这时用户需要等待 VMware vSAN 的重建过程, 包括如何判定一些主机意外宕机所带来的影响。在 VMware vSAN 环境中的操作步骤以及维护方式与传统环境会有比较大的区别。在维护过程中, 如何规范操作而不影响 VMware vSAN 或使 VMware vSAN 重建的收敛时间达到最佳是必须考虑的一个问题。在经过更换 RAID 卡和对 vSAN 环境的操作规范后, VMware vSAN 的重建收敛时间已恢复正常。

4 应用总结

上海公安高等专科学校是国内首个成功部署并上线运行 VMware vSAN 虚拟化存储区域网络的单位, 这对于 VMware vSAN 技术在中国的落地和推广具有重要的示范意义。学校在 VMware vSAN 项目实施过程中遇到的问题以及应对之策对于其他用户的分布式存储实践和 VMware 公司的产品改进都具有积极的意义。上海公安高等专科学校使用 VMware vSphere 服务器虚拟化技术已经有很多年, 至今虚拟机的数量已超过 200 台, 无论从技术先进性、稳定性、安全性还是运维管理的简化上都取得了非常好的效果。学校也希望通过部署 VMware vSAN 虚拟化存储技术, 在存储性能的提升和容量扩展上获得更大收益。

通过 VMware vSAN 技术的成功实施, 上海公安高等专科学校的信息化基础架构得到了进一步优化和提升, 特别是在云服务方面实现了新的跨越。VMware vSAN 给上海公安高等专科学校信息基础架构带来的最大改变在于, 实现了存储的横向扩展和数据安全性的大幅提升。此外, VMware vSAN 还使得学校在信息化基础建设中得到了以下提升:

(1) 让学校在存储建设中摆脱了对于硬件厂商的依赖。
(2) 硬件资源的利用率得到进一步提升, 可以重新评估各种应用部署方案。

(3) VMware vSAN 解决了“知网”等部分大规模数字资源平台不支持 NAS、长期以来只能存放在实体服务器中的问题。

(4)学校的虚拟化平台具有更广泛的适应性,可以保障各种教学云应用的快速部署和上线。

(5)VMware vSAN 的部署为学校存储虚拟化的深入和全面应用以及存储和计算资源的自动化管理奠定了基础。

(6)VMware vSAN 为今后分布式系统的上线提供了技术支持。

(7)VMware vSAN 虚拟化存储特别适合用于大规模的信息化基础平台架构。

从 SLA 的角度来看,vSAN6 不同于传统存储,vSAN6 能够对不同的虚拟机采用不同的存储策略,不同的虚拟机具有不同的服务级别。对于容量管理而言,vSAN6 这种横向扩展存储架构对容量的规划更便利,如果需要增加空间,只需增加相应的磁盘组或者增加服务器即可,并且加入 vSAN 集群的磁盘组会自动对容量进行扩充。对于服务器品牌的兼容性要求而言,只需要满足 VMware 官方提供的兼容性列表即可。而对于传统存储而言,扩大存储容量时需要增加扩展柜,而加入的磁盘无法自动与现有的逻辑卷(LUN)进行自动的融合,必须要管理员进行人为操作。逻辑卷的扩容涉及到多个方面,Raid 磁盘组、Lun 的大小、文件系统大小等,这些操作都会产生安全隐患以及磁盘 IO 的开销。对于性能而言,vSAN 增加磁盘组后得到容量的同时还相应地提升了性能,传统存储添加扩展柜后,存储的控制器处理能力、缓存并没有得到提升,反而控制器需要给更多的磁盘空间进行读写缓存,压力反而增大了。因为传统存储的性能不仅涉及磁盘本身的性能,还涉及控制器的负载能力、控制器与计算节点之间的传输带宽,因此传统存储增加扩展容量后,性能反而会有所下降。

对于未来互联网私有云的平台,在磁盘所需容量无法进行非常准确预估的情况下,vSAN 的横向扩展的架构优势得到了体现,不必为未来可能会达到多少容量而过度地操心,不必为存储过保而操心,不必为是否需要为存储逻辑卷做物理镜像时带来的高资金投入而操心。vSAN 对容量、性能的线性提升,更有利于“一个难以预估、规划使用量的应用”的管理、规划、使用。

结束语 随着云计算、大数据等新技术和应用不断涌现,

传统 SAN 存储在扩展性等方面暴露出不足之处。vSAN 通过软件的不断更新升级,能支持越来越多的功能,性能也得到不断提升其产品发展进程已经完全改变了原来以硬件为主的存储架构演进,随着数据镜像复制、虚拟机双机热备、容灾等高级功能的推出,完全突破了传统 IT 架构,大幅度提高了系统的性能和安全性,IT 管理、数据管理等都将面临变革。上海公安高等专科学校是国内首个成功部署并上线运行 VMware vSAN 虚拟化分布式存储的单位,它在 VMware vSAN 项目实施过程中遇到的问题以及应对之策对于其他用户的分布式存储实践和 VMware 公司的产品改进都具有积极的意义。

参考文献

- [1] 工业和信息化部电信研究院.云计算白皮书[OL]:<http://wenku.baidu.com/view/27a6e43483c4bb4cf7ecd115.html>
- [2] Dean J. Experiences with MapReduce, an abstraction for large-scale computation[C]//International Conference on Parallel Architecture and Compilation Techniques, 2006:1
- [3] Greenberg A, Hamilton JR, Jain N. VL2: A scalable and flexible data center network[C]//Proc. of the SIGCOMM 2009. 2009: 51-62
- [4] Dally W J, Towles B P. Principles and Practices of Interconnection Networks[J]. Macromolecular Materials & Engineering, 2014, 299(6): 707-721
- [5] Abu-Libdeh H, Costa P, Rowstron A, et al. Symbiotic routing in future data centers[C]//Proc. of the SIGCOMM 2010. New Delhi: ACM Press, 2010: 51-62
- [6] Guo C X, Wu H T, Tan K, et al. DCell: A scalable and fault-tolerant network structure for data centers[C]//Proc. of the SIGCOMM 2008. Seattle: ACM Press, 2008: 75-86
- [7] 穆飞, 薛巍. 一种面向大规模存储系统的数据副本映射算法[J]. 计算机研究与发展, 2009, 46(3): 492-497
- [8] 熊润群, 罗军舟. 云计算环境下 QoS 偏好感知的副本选择策略[J]. 通信学报, 2011, 32(7): 93-102
- [9] 王永剑, 裴翔. Nova-BFT: 一种支持多种故障模型的副本状态机协议[J]. 计算机研究与发展, 2011, 48(7): 1134-1145

(上接第 570 页)

结束语 本文基于休假排队,研究顾客在个体均衡和社会最优两方面的进入选择概率,并从附加时间费用角度给出固定和可变进入费用制定原则。与之前顾客行为研究的不同点在于,文中分析了服务前收费(事前费用)和服务后收费(事后收费)两种费用方式,并引入了附加时间效用函数来体现顾客费用,为进一步研究其它排队行为提出了新的角度和思想。

参考文献

- [1] Naor P. The regulation of queue size by levying tolls[J]. Econometrica, 1969, 37(1): 15-24
- [2] Haviv M, Kerner Y. On balking from an empty queue [J]. Queueing System, 2007, 55(55): 239-249
- [3] Economou A, Kanta S. Equilibrium customer strategies and social-profit maximization in the single-server constant retrial

queue[J]. Naval Research Logistics, 2011, 58(2): 107-122

- [4] Haviv M. Regulating an M/G/1 queue when customers know their demand[J]. Performance Evaluation, 2014, 77: 57-71
- [5] Hassin R, Haviv M. To Queue or Not to Queue; Equilibrium Behavior in Queueing Systems [M]. Boston: Kluwer Academic Publishers, 2003
- [6] Sun W, Wang Y, Tian N. Pricing and setup/closedown policies in unobservable queues with strategic customers[J]. 4or Quarterly Journal of the Belgian French & Italian Operations Research Societies, 2012, 10(10): 287-311
- [7] Zhang F, Wang J, Liu B. Strategic joining in M/M/1 retrial queue[J]. European Journal of Operational Research, 2013, 230(1): 76-87
- [8] Tian N, Zhang G Z. Vacation queueing models-theory and applications [M]. New York: Springer-Verlag, 2006